Section Outline

1. Problems (with Solutions)

Problem 1: Crystal Growth
(based on a problem from the text)

Background: Crystalline forms of certain chemical compounds are used in various electronic devices, and it is often more desirable to have large crystals rather than small ones. Crystals of one particular compound are to be produced by a commercial process and an investigator wants to examine relationship between the size of the crystal, determined by its weight in grams, and the number of hours it takes the crystal to grow to its final size. The following data are from a laboratory study in which 14 crystals of various sizes were obtained by allowing the crystals to grow for different preselected amounts of time.

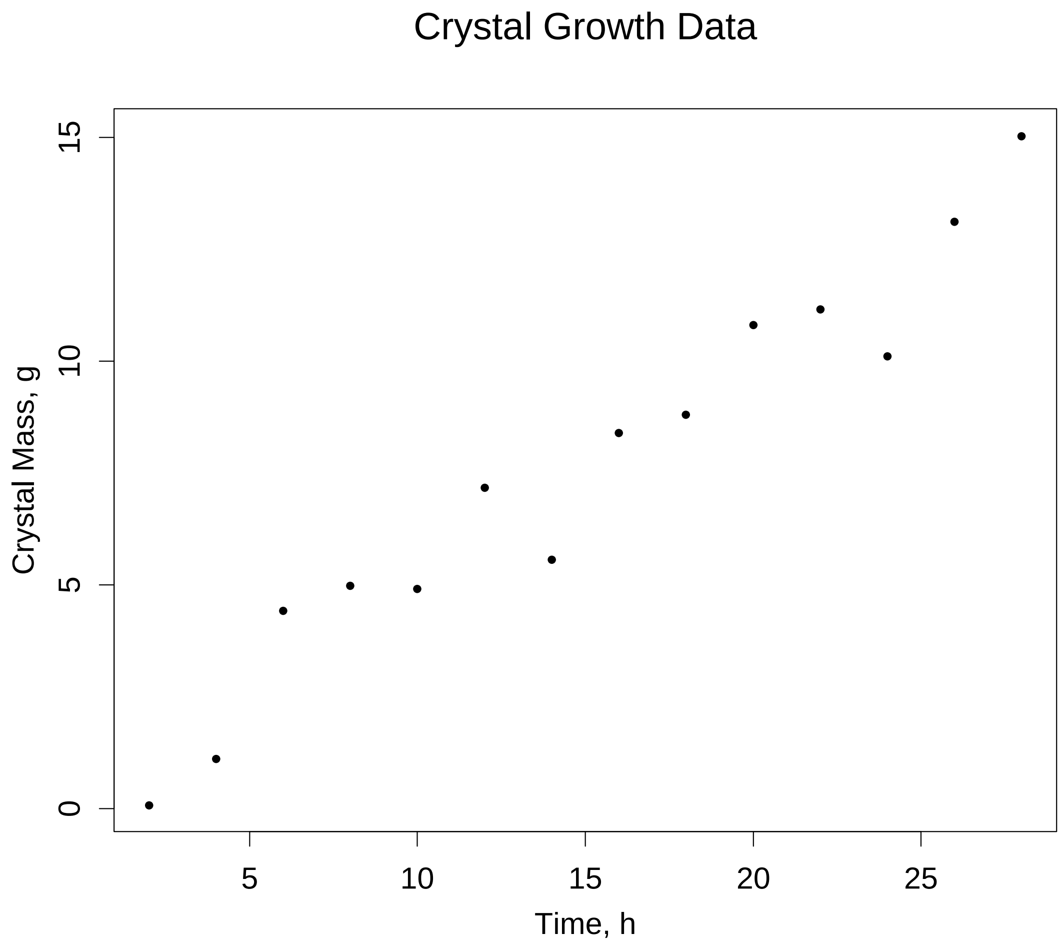
Weight (grams)	Time (hours)
0.08	2
1.12	4
4.43	6
4.98	8
4.92	10
7.18	12
5.57	14
8.40	16
8.81	18
10.81	20
11.16	22
10.12	24
13.12	26
15.04	28

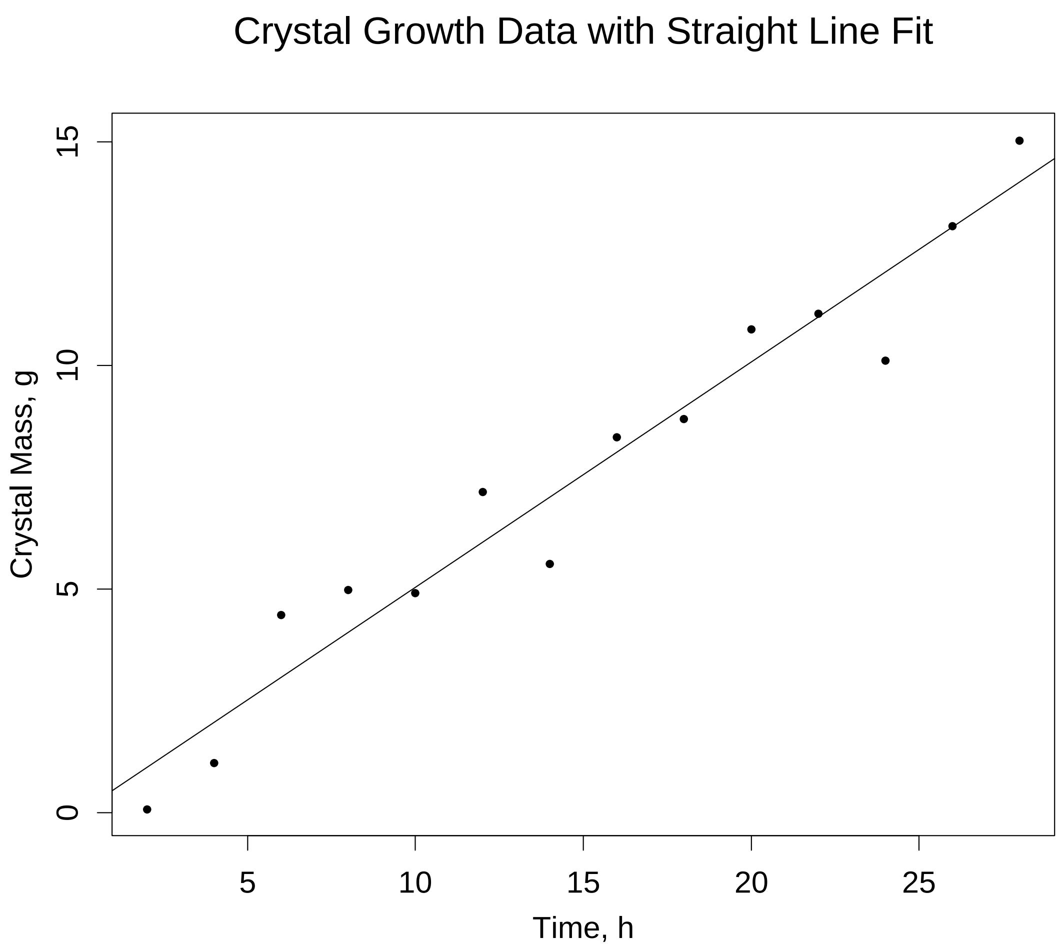
1a. Fit and validate a model appropriate to the data.

1b. If a crystal is grown for 15 hours what is its predicted weight?

1c. Compute an interval that will contain the weight of the crystal with 95% confidence.

1d. For this process to be comercially viable suppose matched sets of six crystals will be grown for 24 hours in a sealed growth chamber under environmental conditions that are difficult to establish in the chamber. If every crystal in a set must end up within 2 g of the nominal weight of 12 grams with a probability of 0.9, does this process look viable? If not, could the apparent non-viability of the process be an artifact of the experiment design?


Crystal Growth Data
Crystal Mass, g
0
5
10
15
5
10
15
20
25
Time, h

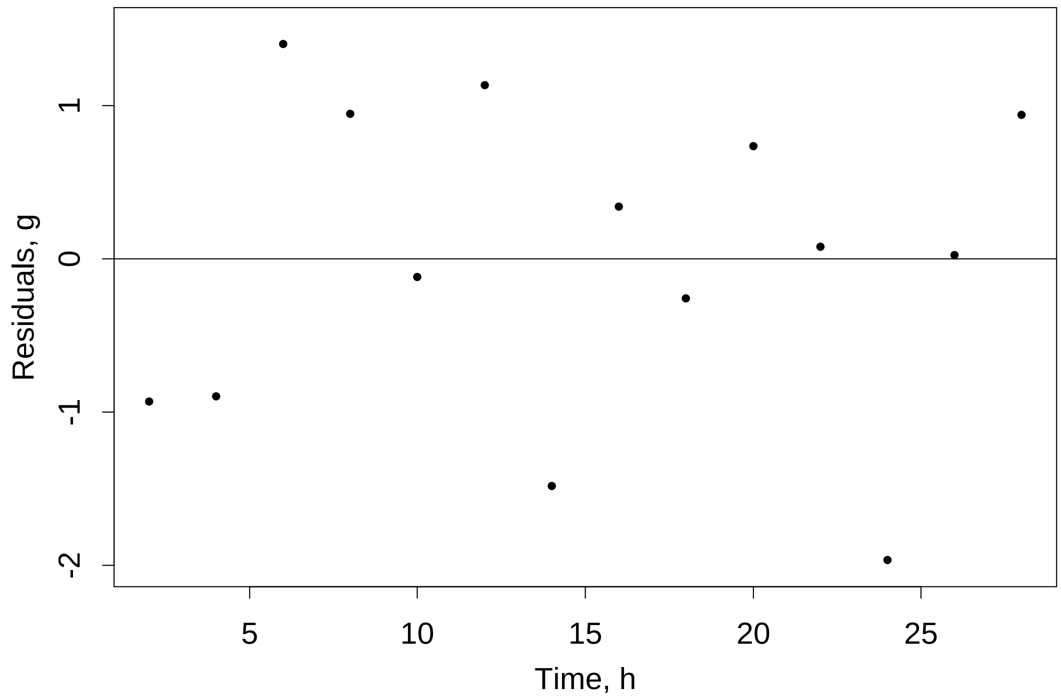

Crystal Growth Data with Straight Line Fit
Crystal Mass, g
0
5
10
15
5
10
15
20
25
Time, h

Data, Residuals, and Other Diagnostic Values

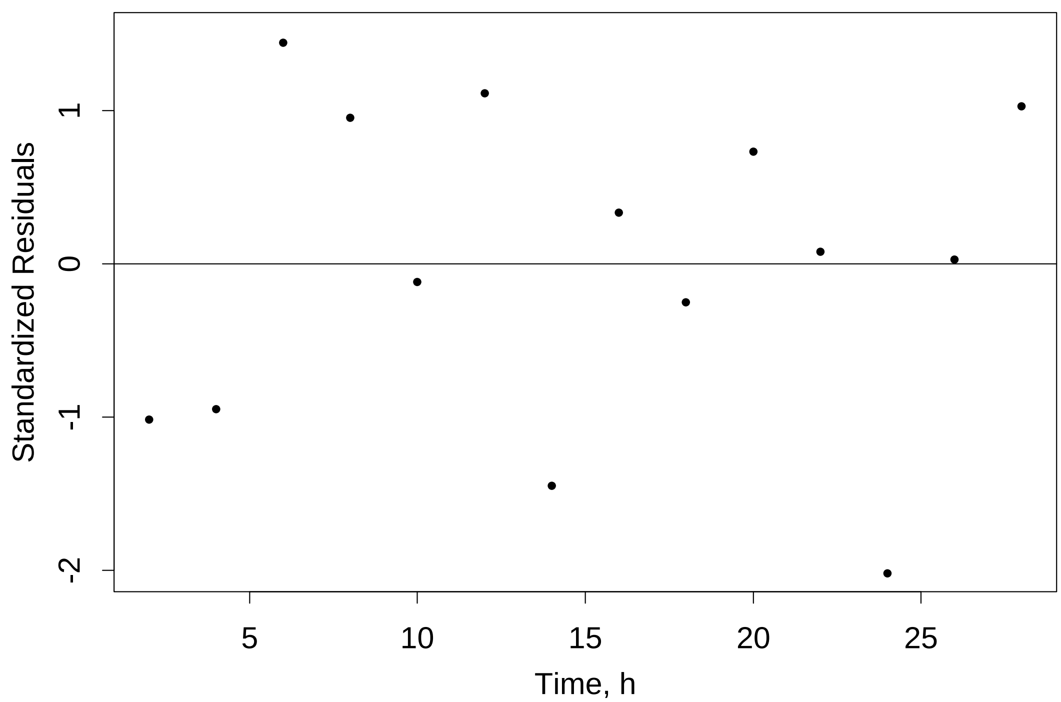
	Time	Mass	Raw Residuals	Standardized Residuals	Studentized Residuals
1	2	0.08	-0.92828571	-1.01437798	-1.01571604
2	4	1.12	-0.89514286	-0.94518009	-0.94063181
3	6	4.43	1.40800000	1.44726304	1.52513102
4	8	4.98	0.95114286	0.95781297	0.95423719
5	10	4.92	-0.11571429	-0.11480784	-0.10998056
6	12	7.18	1.13742857	1.11766716	1.13054547
7	14	5.57	-1.47942857	-1.44681923	-1.52456449
8	16	8.40	0.34371429	0.33613819	0.32335372
9	18	8.81	-0.25314286	-0.24874481	-0.23877140
10	20	10.81	0.74000000	0.73420323	0.71928732
11	22	11.16	0.08314286	0.08372592	0.08018489
12	24	10.12	-1.96371429	-2.01847379	-2.37792910
13	26	13.12	0.02942857	0.03107359	0.02975189
14	28	15.04	0.94257143	1.02998860	1.03285058

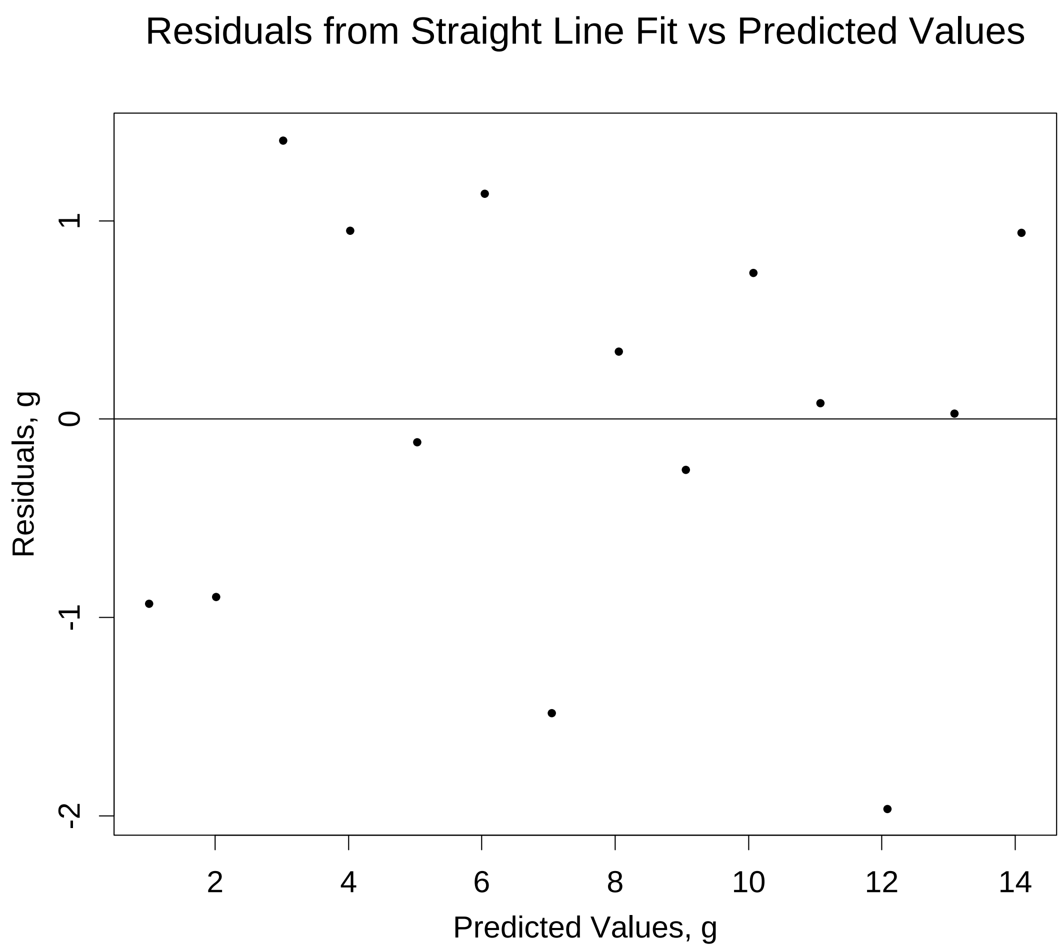
	Time	Mass	Cook's Distance	DFFITS
1	2	0.08	0.1780896965	-0.59759491
2	4	1.12	0.1147554994	-0.47676773
3	6	4.43	0.2001356438	0.66670990
4	8	4.98	0.0656936227	0.36112057
5	10	4.92	0.0007233388	-0.03643596
6	12	7.18	0.0552866691	0.33635722
7	14	5.57	0.0818464864	-0.42633064
8	16	8.40	0.0044178117	0.09042294
9	18	8.81	0.0027384418	-0.07103870
10	20	10.81	0.0295822524	0.23829599
11	22	11.16	0.0005019744	0.03034509
12	24	10.12	0.3892922269	-1.03950995
13	26	13.12	0.0001240301	0.01508002
14	28	15.04	0.1836132434	0.60767599

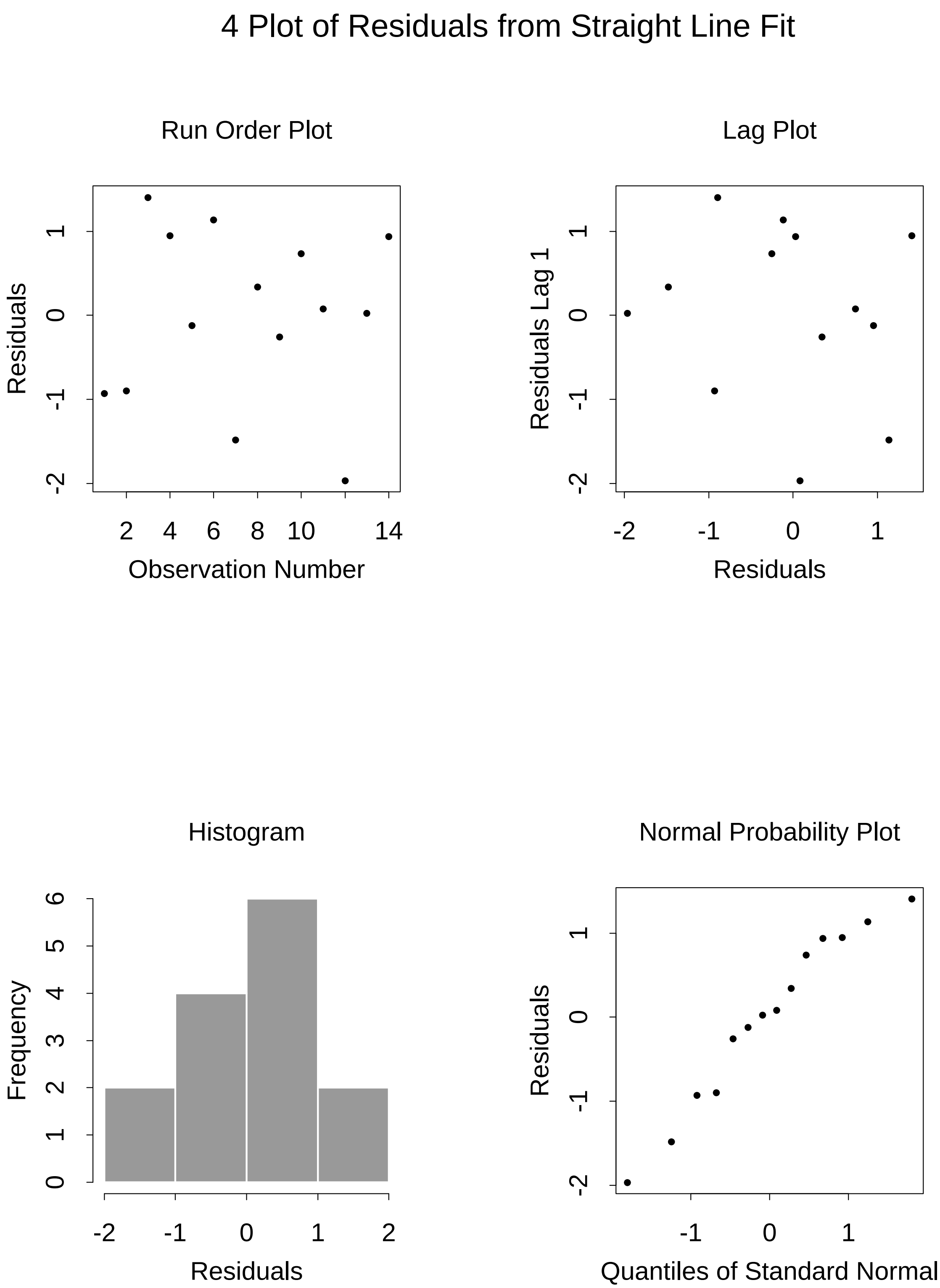
Raw Residuals from Straight Line Fit vs Time

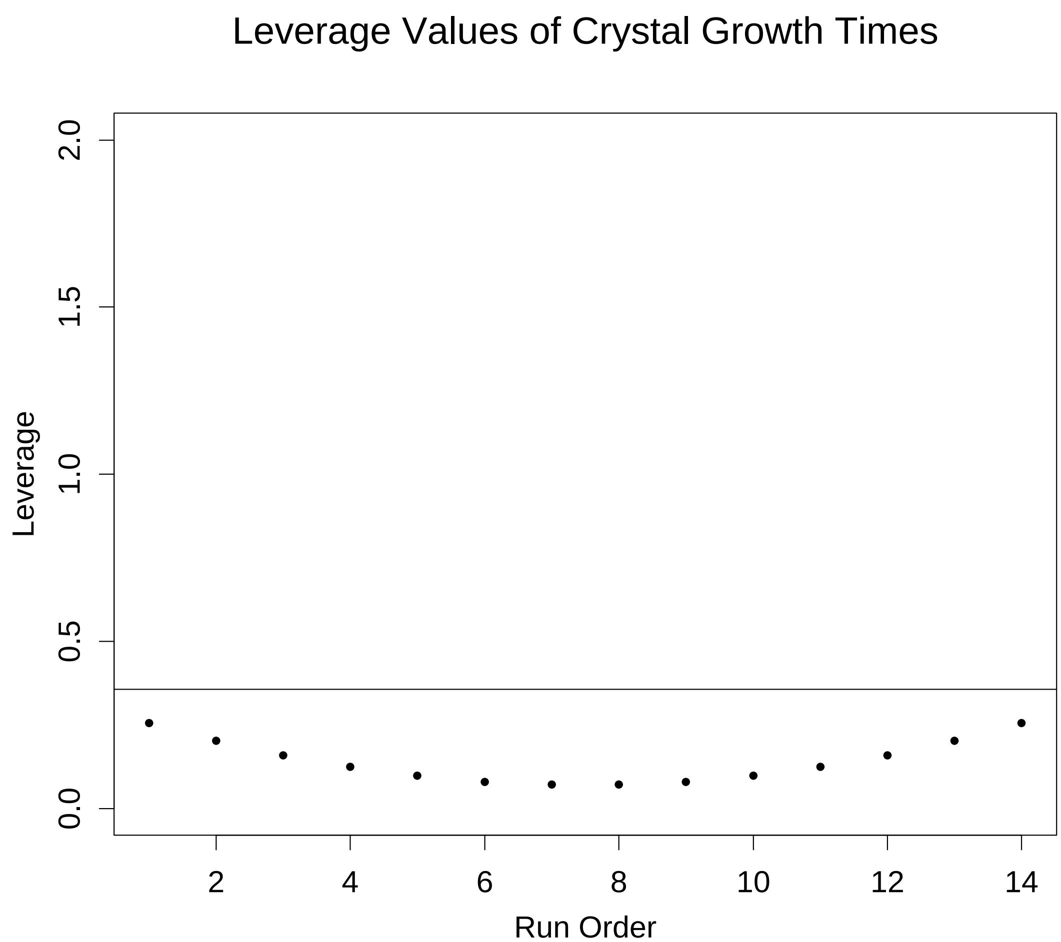


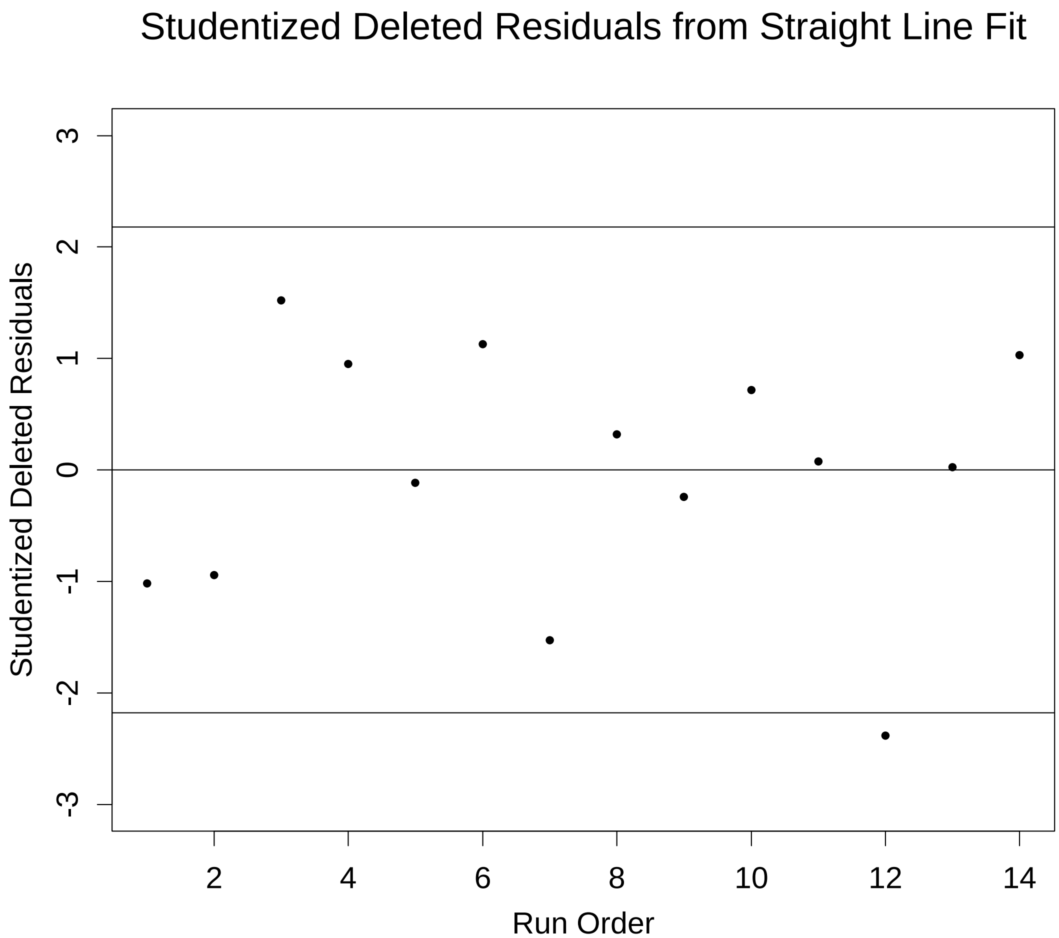
Standardized Residuals from Straight Line Fit vs Time

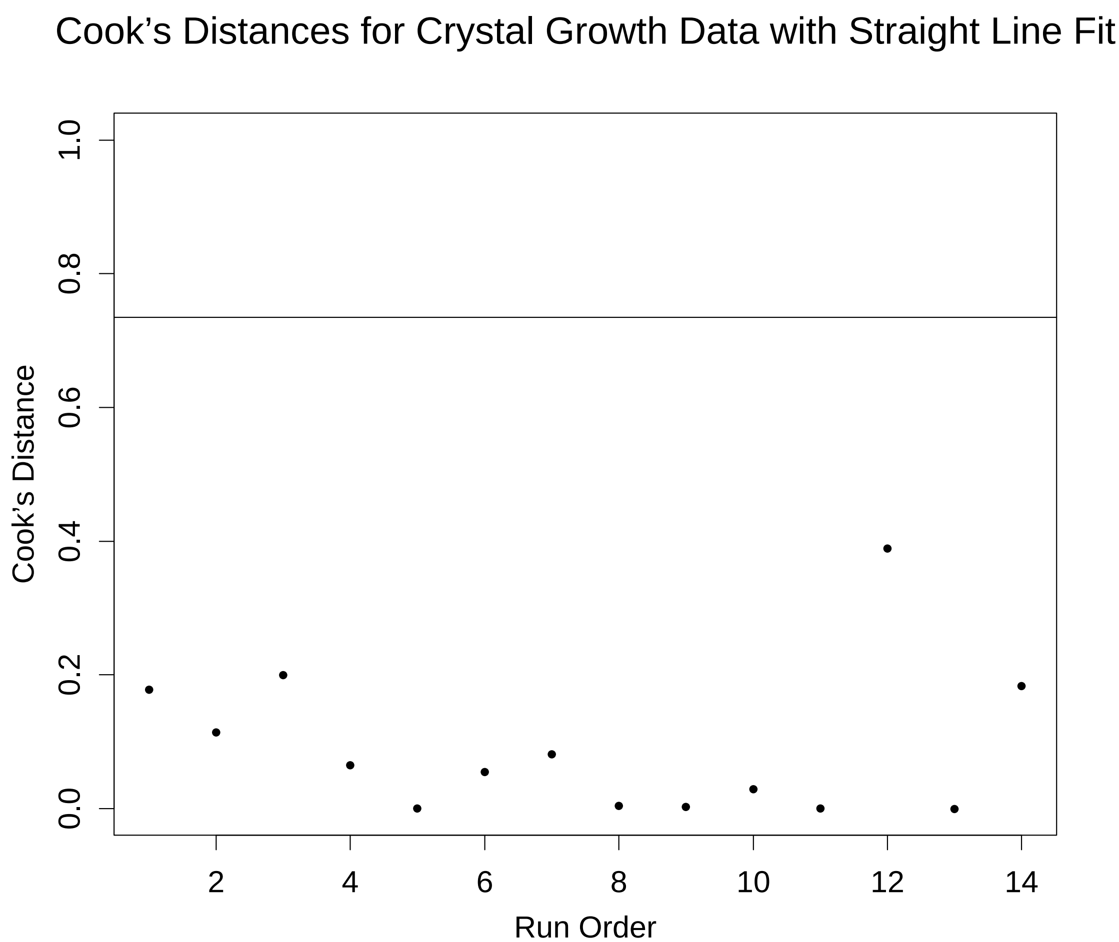


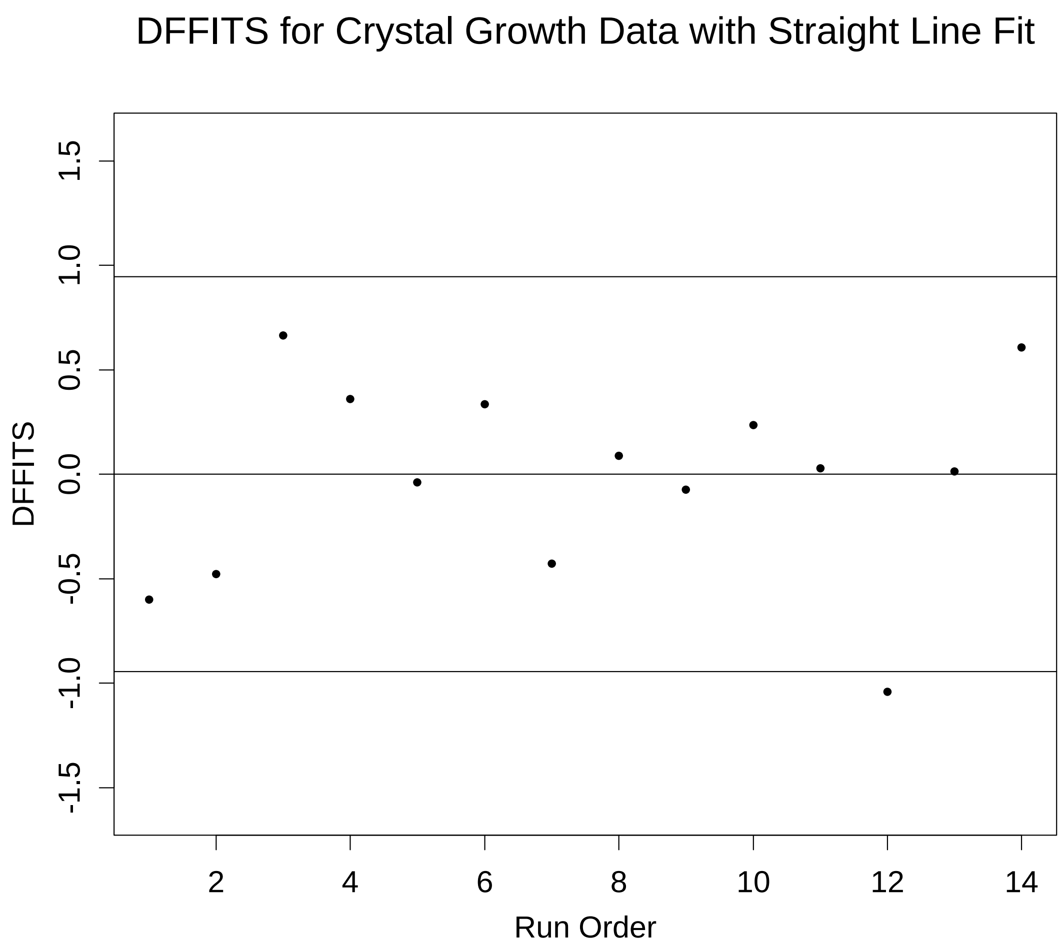

Residuals from Straight Line Fit vs Predicted Values
Residuals, g
1
0
-1
-2
2
4
6
8
10
12
14
Predicted Values, g


4 Plot of Residuals from Straight Line Fit
Run Order Plot
Residuals
Observation Number
Lag Plot
Residuals Lag 1
Residuals
Histogram
Frequency
Residuals
Normal Probability Plot
Residuals
Quantiles of Standard Normal

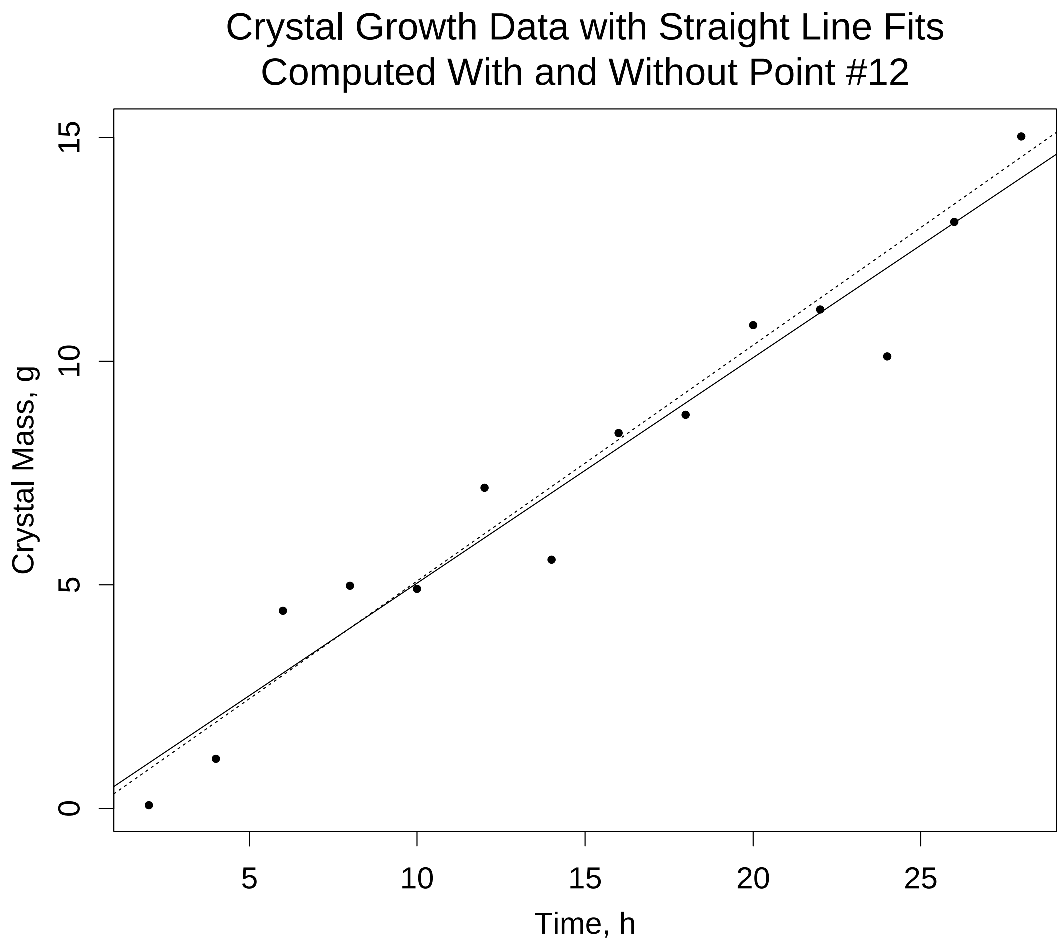

Leverage Values of Crystal Growth Times
Leverage
0.0
0.5
1.0
1.5
2.0
2
4
6
8
10
12
14
Run Order


Studentized Deleted Residuals from Straight Line Fit
Studentized Deleted Residuals
Run Order
3
2
1
0
-1
-2
-3
2
4
6
8
10
12
14


Cook's Distances for Crystal Growth Data with Straight Line Fit
Cook's Distance
Run Order
0.0
0.2
0.4
0.6
0.8
1.0
2
4
6
8
10
12
14



DFFITS for Crystal Growth Data with Straight Line Fit
DFFITS
Run Order
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
2
4
6
8
10
12
14


Crystal Growth Data with Straight Line Fits
Computed With and Without Point #12
Crystal Mass, g
0
5
10
15
5
10
15
20
25
Time, h

```
Output from Straight Line Fit - All Data

N = 14

Residual Standard Error = 1.0618

Multiple R-Square = 0.9446

F-statistic = 204.5779 on 1 and 12 df, p-value = 0

            coef std.err  t.stat p.value
Intercept 0.0014  0.5994  0.0024  0.9981
     Time 0.5034  0.0352 14.3031  0.0000
```

```
Output from Straight Line Fit - Point #12 Removed

N = 13

Residual Standard Error = 0.9013

Multiple R-Square = 0.9623

F-statistic = 280.8466 on 1 and 11 df, p-value = 0

             coef std.err   t.stat p.value
Intercept -0.1785  0.5144  -0.3470  0.7351
     Time  0.5266  0.0314  16.7585  0.0000
```

Individual 95% Prediction Interval

Plugging in the numbers from the regression output yields:

Model 1: All Data

$$\hat{y}^* = 0.0014 + 0.5034 \times 15 = 7.5524$$

$$\begin{aligned} U &= 2.178813(1.0618)\sqrt{1 + \frac{1}{14} + \frac{(15-15)^2}{910}} \\ &= 2.2395 \end{aligned}$$

Model 2: Point #12 Removed

$$\hat{y}^* = -0.1785 + 0.5266 \times 15 = 7.7205$$

$$\begin{aligned} U &= 2.200985(0.9013)\sqrt{1 + \frac{1}{13} + \frac{(15-14.30769)^2}{840}} \\ &= 2.1374 \end{aligned}$$

Simultaneous 90% Prediction Intervals

Plugging in the numbers from the regression output yields:

Model 1: All Data

$$\hat{y}^* = 0.0014 + 0.5034 \times 24 = 12.083$$

$$\begin{aligned} U &= 2.779473(1.0618)\sqrt{1 + \frac{1}{14} + \frac{(24-15)^2}{910}} \\ &= 3.179 \end{aligned}$$

Model 2: Point #12 Removed

$$\hat{y}^* = -0.1785 + 0.5266 \times 24 = 12.4599$$

$$\begin{aligned} U &= 2.820034(0.9013)\sqrt{1 + \frac{1}{13} + \frac{(24-14.30769)^2}{840}} \\ &= 2.771 \end{aligned}$$

Computations of Coverage Factors for the Simultaneous Prediction Intervals

The coverage factor for the simultaneous intervals for Model 1 is chosen to be the lesser of:

$$t_{1-0.1/(2(6)),12} = 2.779473$$

and

$$\sqrt{6F_{0.9,6,12}} = 3.739805$$

The coverage factor for the simultaneous intervals for Model 2 is chosen to be the lesser of:

$$t_{1-0.1/(2(6)),11} = 2.820034$$

and

$$\sqrt{6F_{0.9,6,11}} = 3.78608$$

Problem 2: Measurement of Arsenic in Water
(based on a problem from the text)

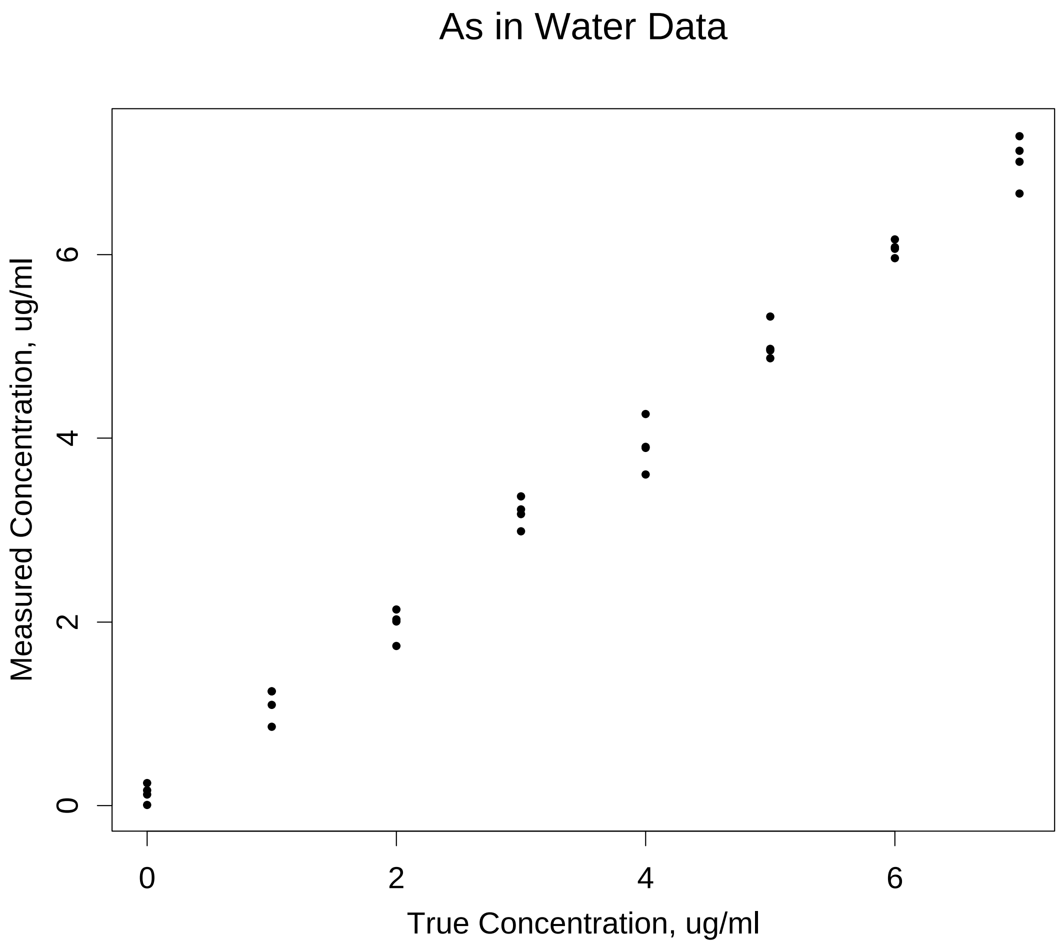
Background: An investigator wants to evaluate the performance of a new laboratory method for analyzing the concentration of arsenic (As) in water samples that is much cheaper than the existing method. If the new method is proven to be scientifically acceptable it will be adopted by environmental research groups for monitoring the quality of As in industrial waste water. To investigate the relationship between the measured concentrations of As and the actual concentrations, the investigator makes several water samples containing known (preselected) amounts of As. These water samples are analyzed by a laboratory technician (who is unaware of the actual amount of As in these solutions) using the new method of analysis. The concentrations are reported in ug/ml. The data are listed below.

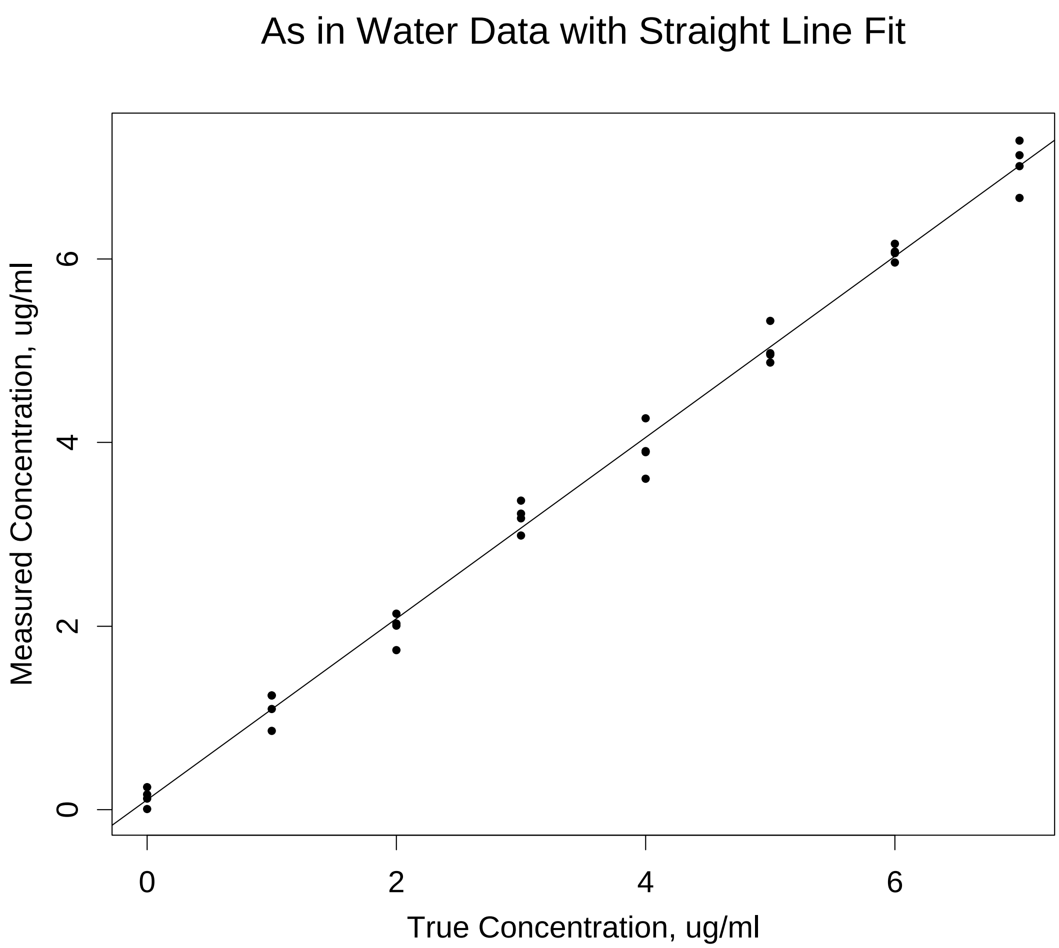
Measured Concentration (ug/ml)	True Concentration (ug/ml)	Measured Concentration (ug/ml)	True Concentration (ug/ml)
0.17	0	3.91	4
0.25	0	3.90	4
0.01	0	3.61	4
0.12	0	4.27	4
1.25	1	4.88	5
0.86	1	5.33	5
1.25	1	4.96	5
1.10	1	4.98	5
2.01	2	6.09	6
2.03	2	6.17	6
2.14	2	6.07	6
1.74	2	5.97	6
3.18	3	6.67	7
2.99	3	7.02	7
3.23	3	7.14	7
3.37	3	7.30	7

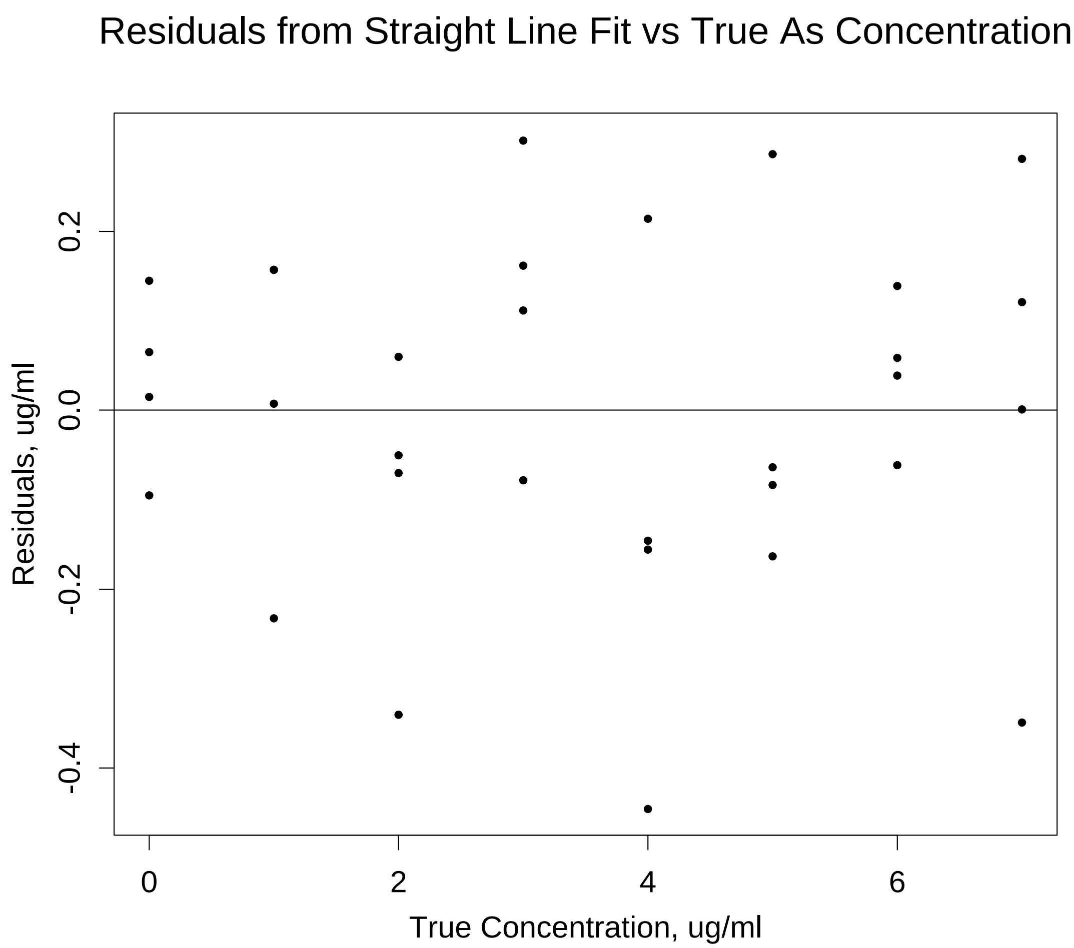
2a. Fit and validate a model appropriate to the data.

2b. What is the estimated concentration of As in a solution which gives a measurement result of 1.32 ug/ml? What is the expanded uncertainty of this result at the 95% confidence level?

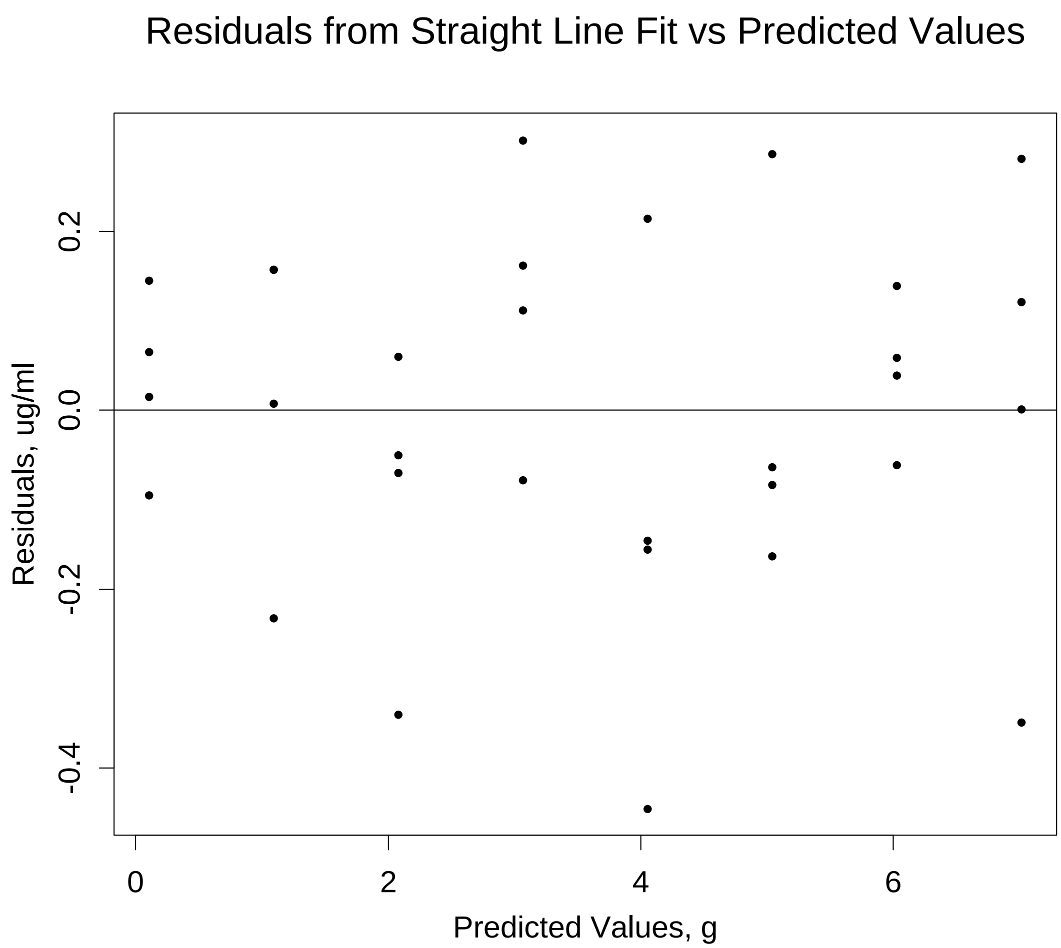
2c. Compute a 95% confidence interval for the true concentration of As in a solution which gives a measurement result of 8.87 ug/ml.

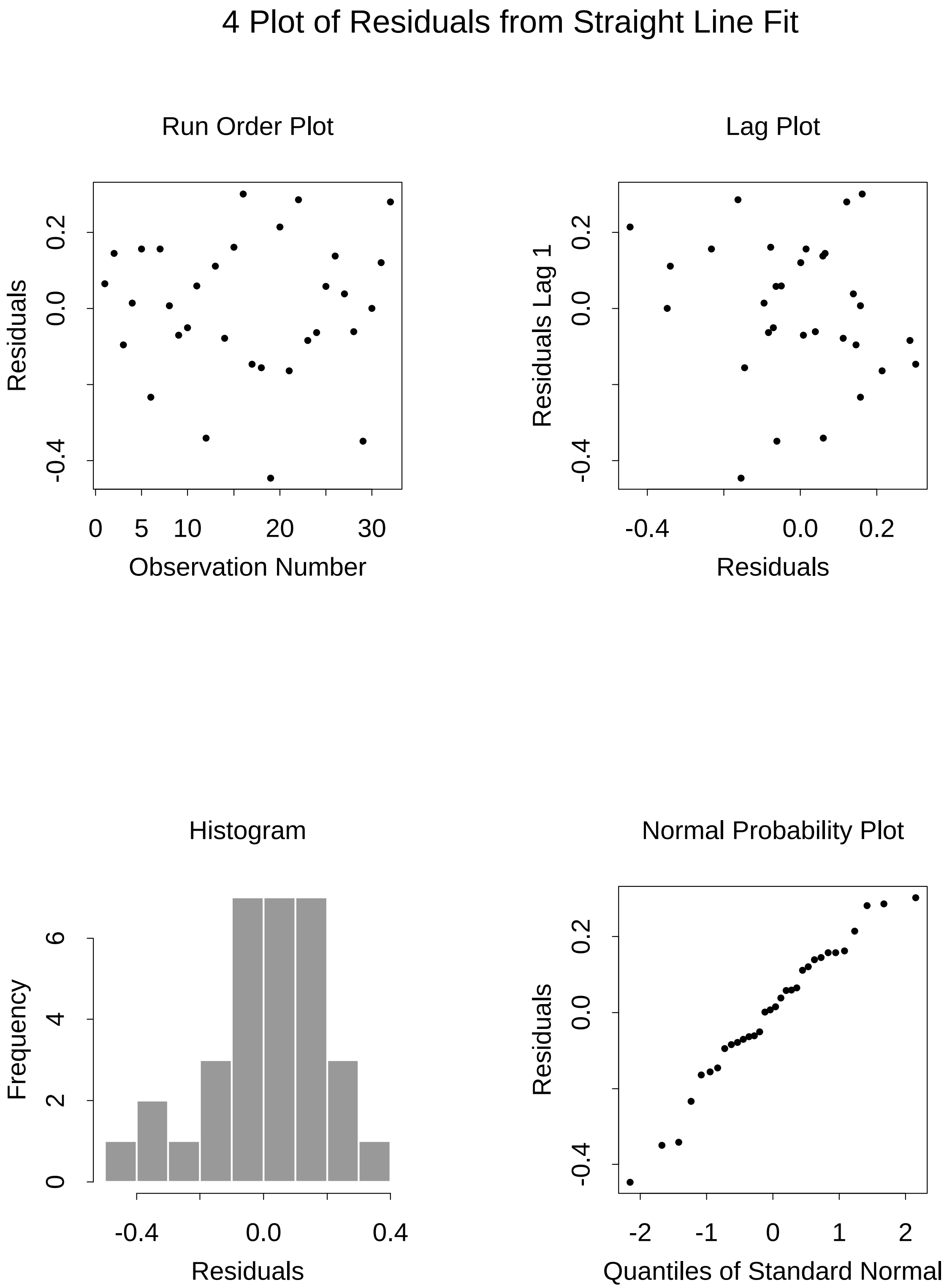

As in Water Data
Measured Concentration, ug/ml
True Concentration, ug/ml
0
2
4
6
0
2
4
6

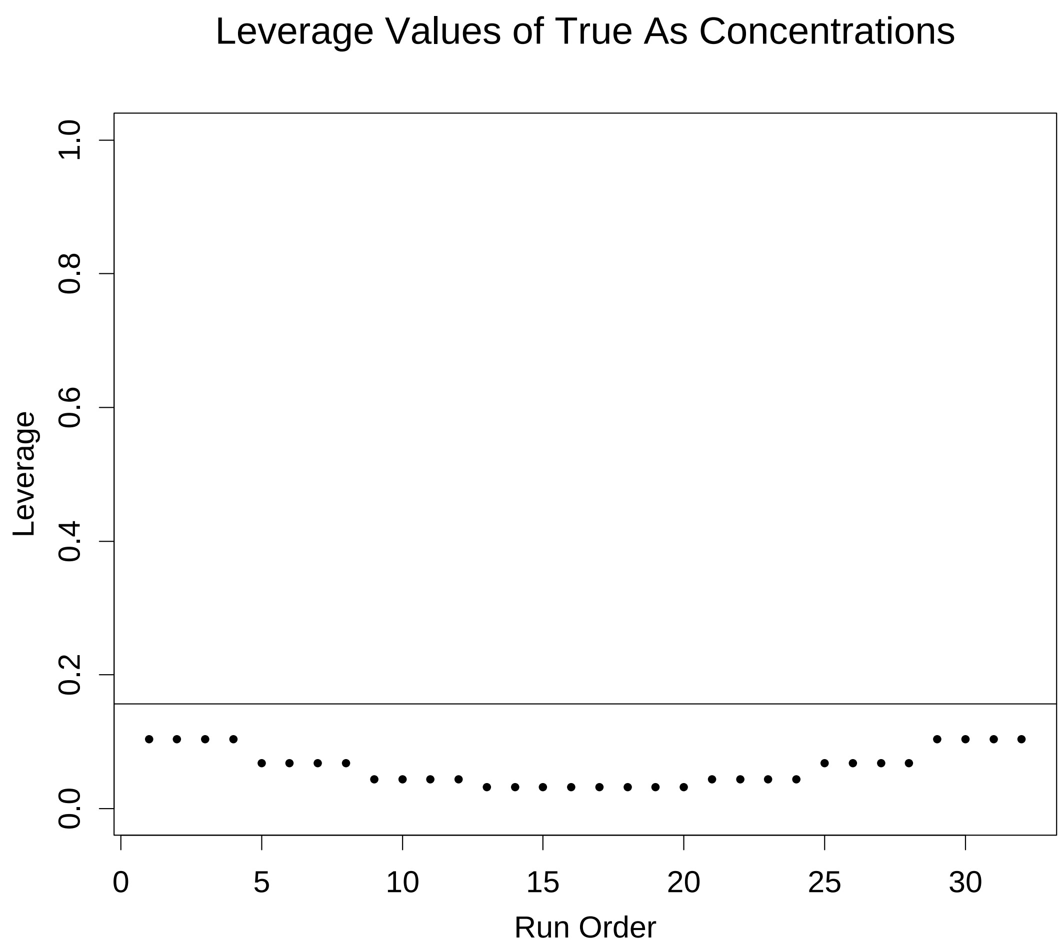

As in Water Data with Straight Line Fit
Measured Concentration, ug/ml
0
2
4
6
0
2
4
6
True Concentration, ug/ml

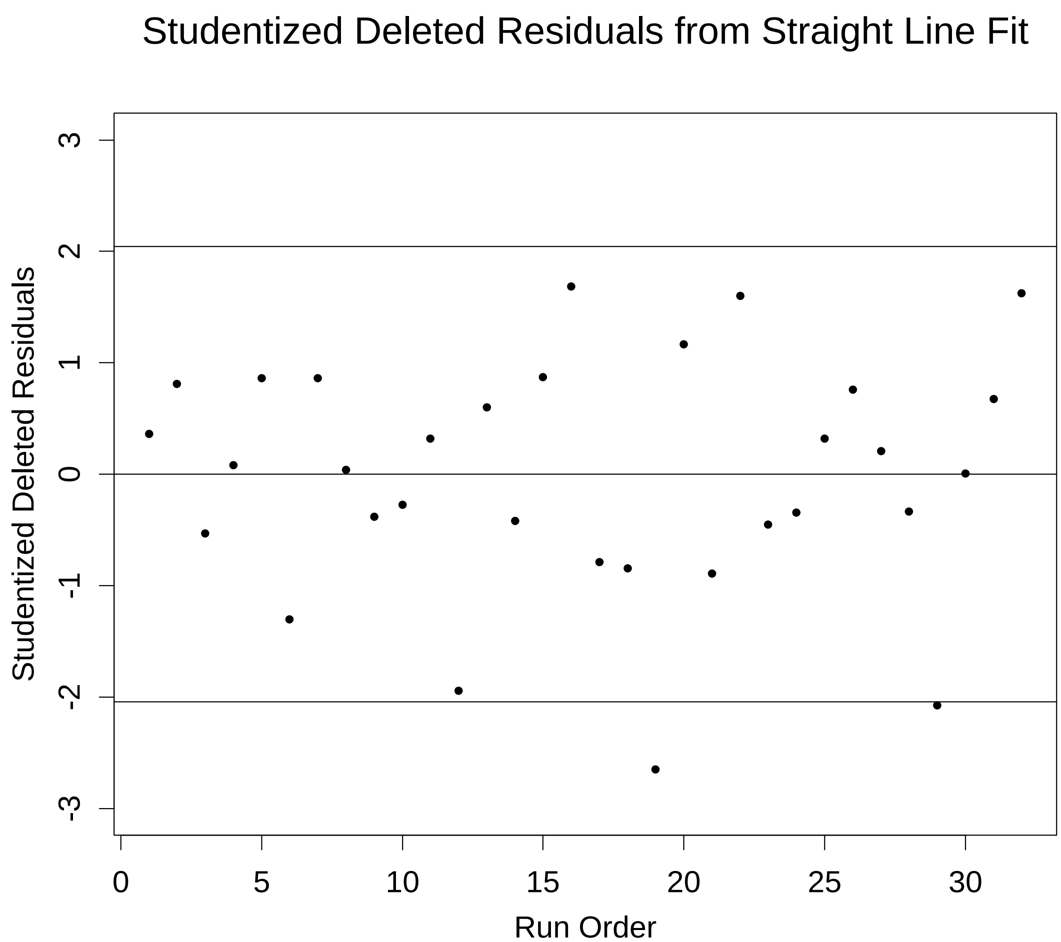


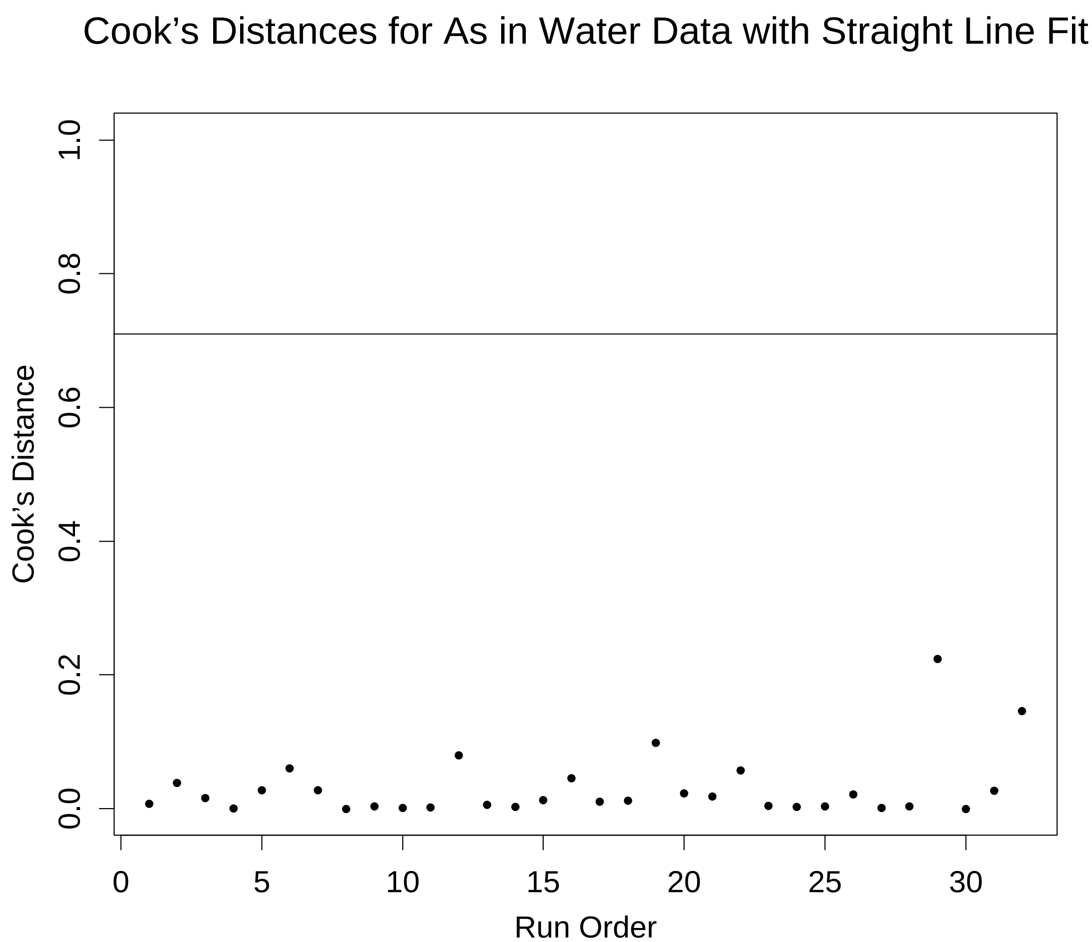
Residuals from Straight Line Fit vs True As Concentration
Residuals, ug/ml
0.2
0.0
-0.2
-0.4
0
2
4
6
True Concentration, ug/ml


Residuals from Straight Line Fit vs Predicted Values
Residuals, ug/ml
0.2
0.0
-0.2
-0.4
0
2
4
6
Predicted Values, g

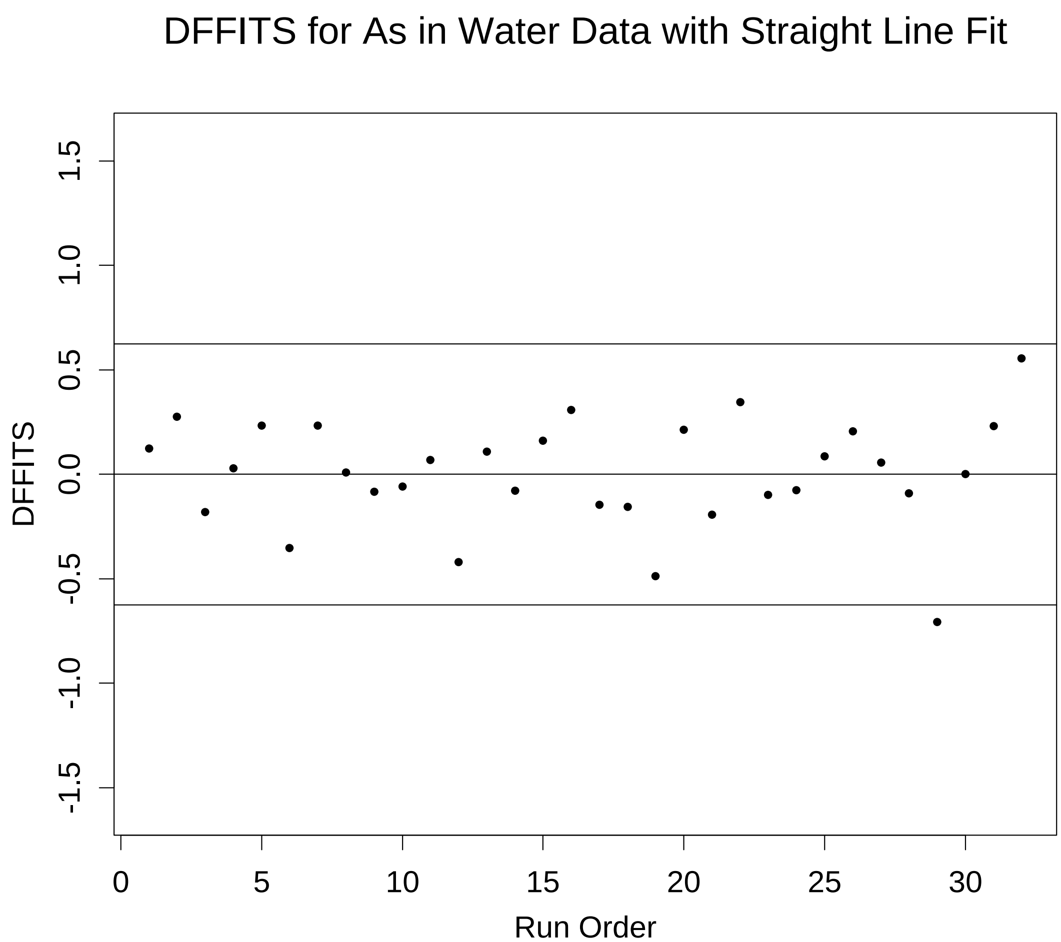

4 Plot of Residuals from Straight Line Fit
Run Order Plot
Residuals
Observation Number
Lag Plot
Residuals Lag 1
Residuals
Histogram
Frequency
Residuals
Normal Probability Plot
Residuals
Quantiles of Standard Normal

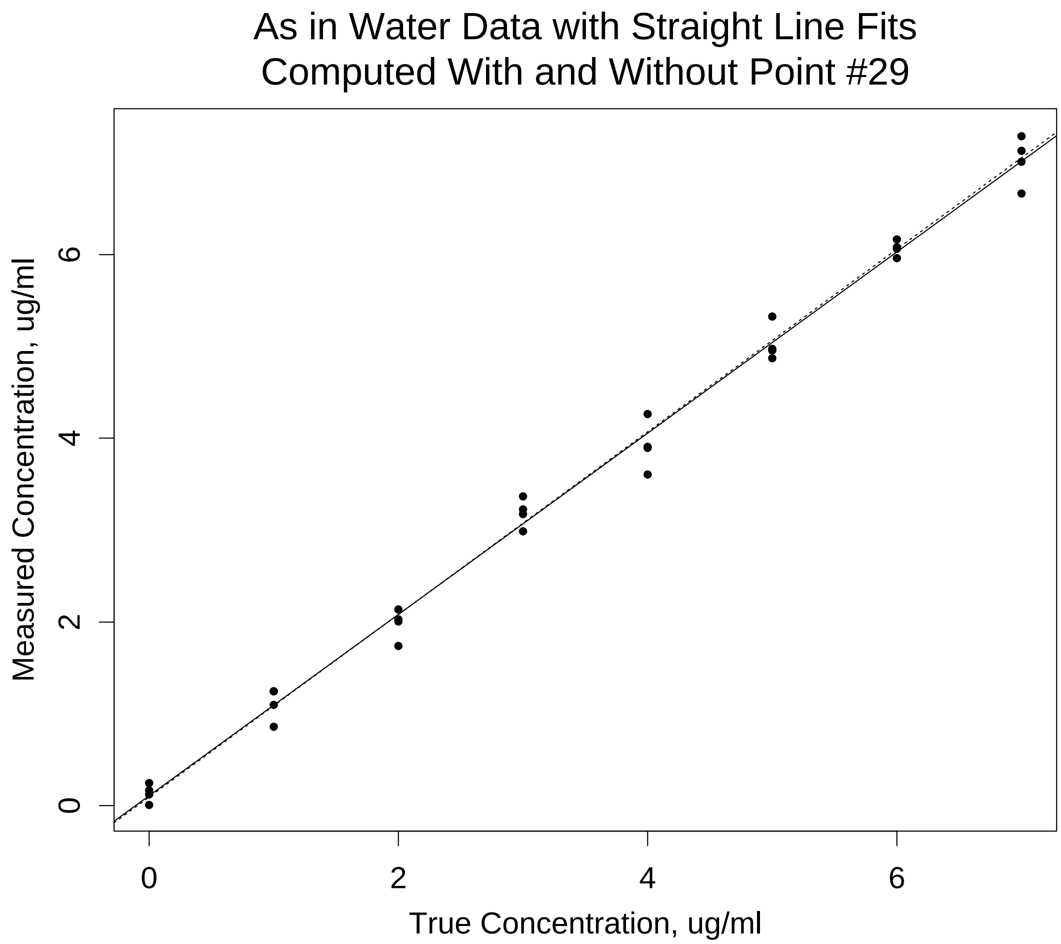

Leverage Values of True As Concentrations
Leverage
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
Run Order


Studentized Deleted Residuals from Straight Line Fit
Studentized Deleted Residuals
-3
-2
-1
0
1
2
3
0
5
10
15
20
25
30
Run Order



Cook's Distances for As in Water Data with Straight Line Fit
Cook's Distance
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
Run Order


DFFITS for As in Water Data with Straight Line Fit
DFFITS
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
0
5
10
15
20
25
30
Run Order


As in Water Data with Straight Line Fits
Computed With and Without Point #29
Measured Concentration, ug/ml
0
2
4
6
True Concentration, ug/ml
0
2
4
6

```
Output from Straight Line Fit - All Data

N = 32

Residual Standard Error = 0.1875

Multiple R-Square = 0.9936

F-statistic = 4663.009 on 1 and 30 df, p-value = 0

            coef std.err  t.stat p.value
Intercept 0.1046  0.0605  1.7284  0.0942
True Conc 0.9877  0.0145 68.2862  0.0000
```

```
Output from Straight Line Fit - With Point #29 Removed

N = 31

Residual Standard Error = 0.178

Multiple R-Square = 0.9941

F-statistic = 4862.312 on 1 and 29 df, p-value = 0

            coef std.err  t.stat p.value
Intercept 0.0884  0.0580  1.5242  0.1383
True Conc 0.9958  0.0143 69.7303  0.0000
```

95% Calibration Interval Using All Data

$$\hat{x}^* = \frac{1.32 - 0.1046}{0.9877} = 1.2305$$

$$\begin{aligned} U &= 2.042272\sqrt{\frac{0.1875^2}{0.9877^2}(1 + \frac{1}{32} + \frac{(1.2305 - 3.5)}{168}} \\ &= 0.3995 \end{aligned}$$

$$c = \frac{2.042272^2 \times 0.1875^2}{0.9877^2 \times 168} \approx 0.0009$$

95% Calibration Interval Using All Data

$$\hat{x}^* = \frac{8.87 - 0.1046}{0.9877} = 8.8746$$

$$U = 2.042272\sqrt{\frac{0.1875^2}{0.9877^2}(1 + \frac{1}{32} + \frac{(8.8746 - 3.5)}{168}}$$

$$= 0.4253$$

$$c = \frac{2.042272^2 \times 0.1875^2}{0.9877^2 \times 168} \approx 0.0009$$

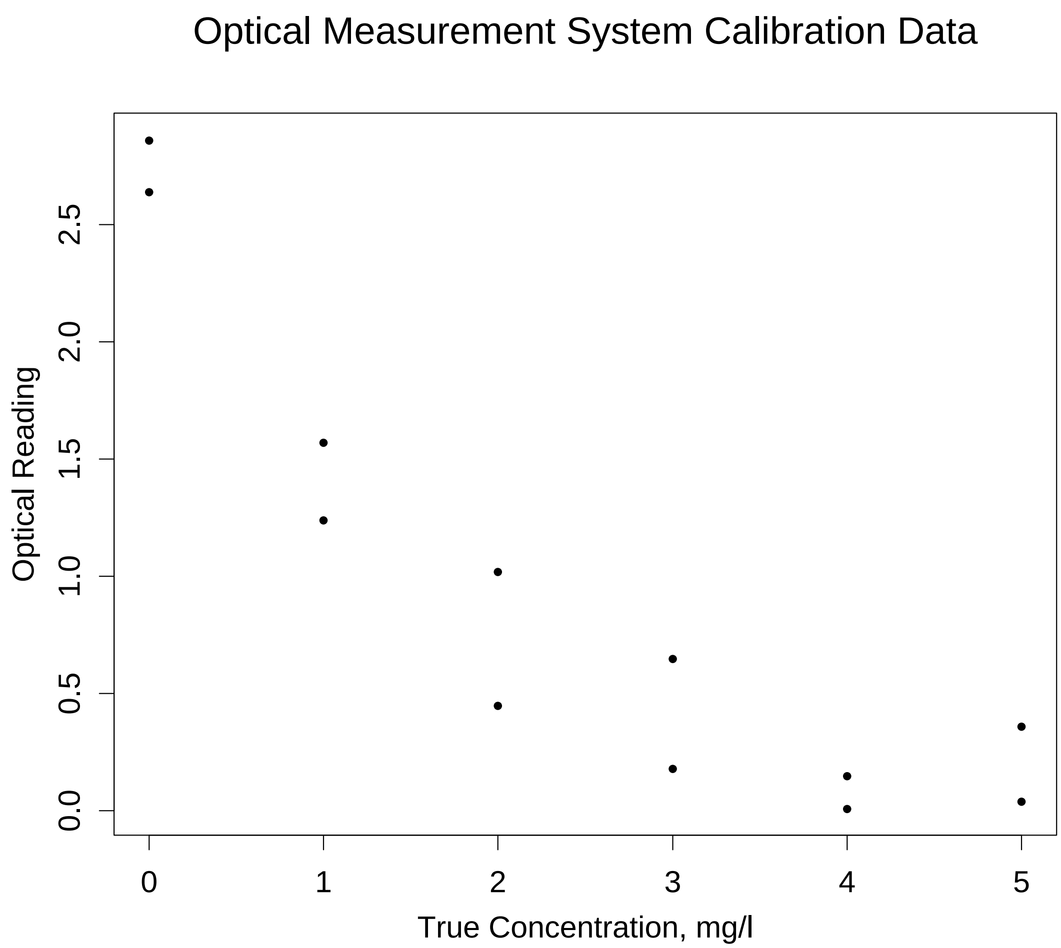
Problem 3: Calibration of an Optical Measurement Method
(based on a problem from the text)

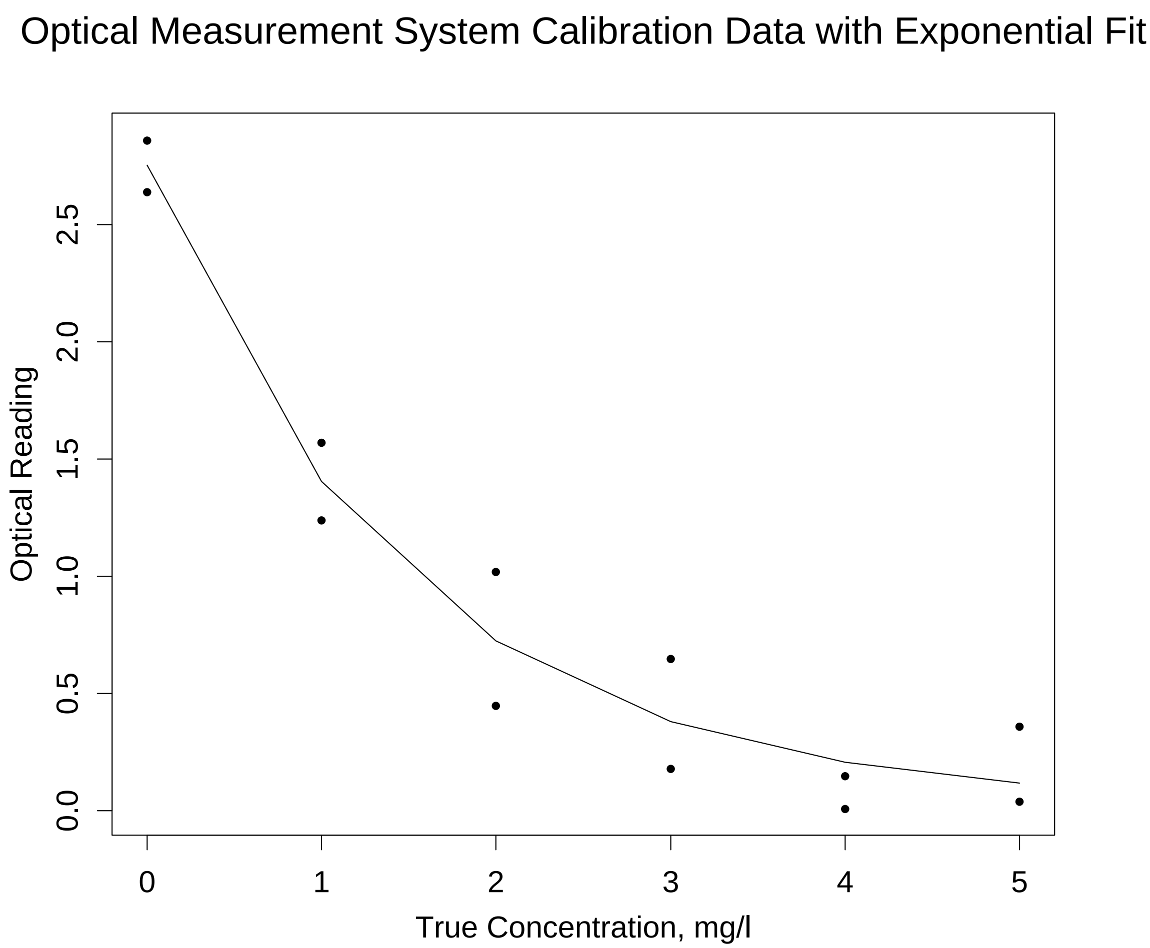
Background: It is well known that when a beam of light is passed through a chemical solution a certain fraction of the indcident light will be absorbed or reflected and the remainder will be transmitted. The intensity of the transmitted light decreases as the concentration of the chemical solution increases. This fact is often used to determine concentrations of various chemicals in solution. The data listed below are from an experiment in which several solutions of known concentrations of a pure chemical were used to measure the amount of transmitted light to determine the relationship between the optical readings and the true concentrations of the chemical.

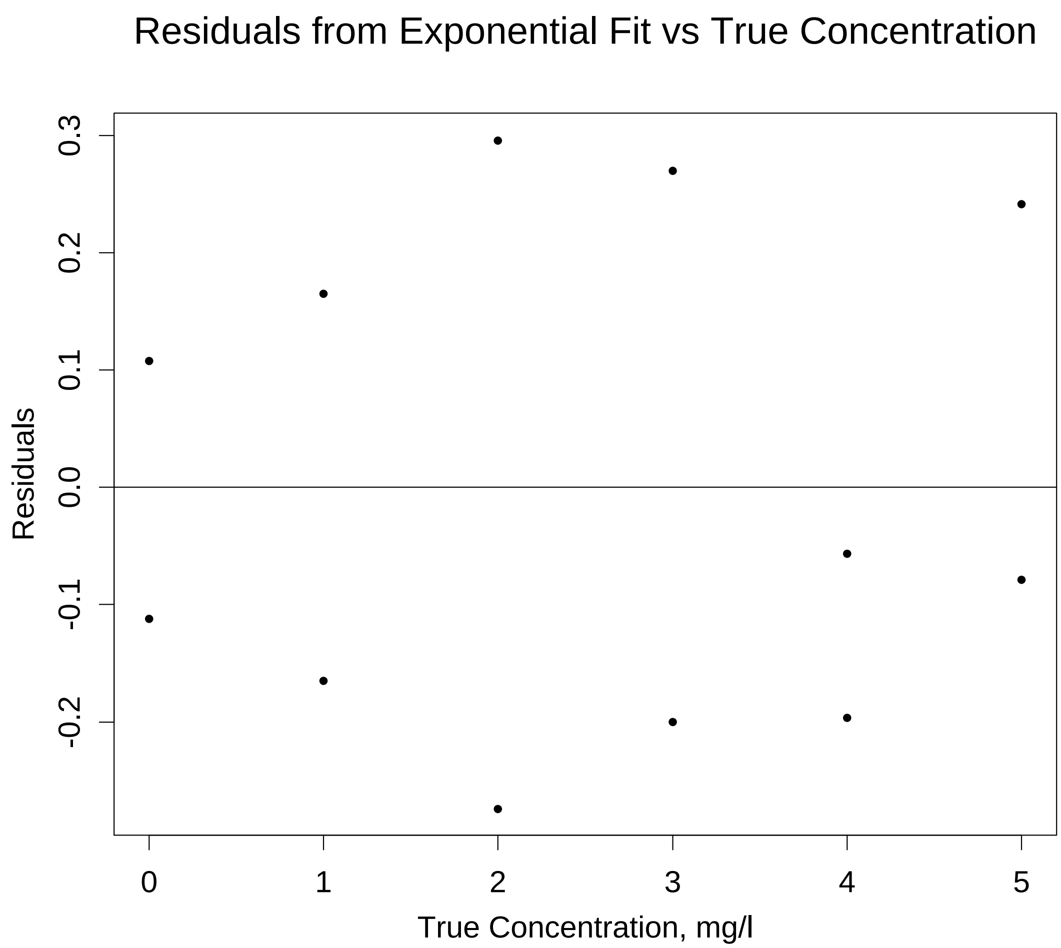
Optical Reading (relative units)	True Concentration (mg/l)
2.86	0
2.64	0
1.57	1
1.24	1
0.45	2
1.02	2
0.65	3
0.18	3
0.15	4
0.01	4
0.04	5
0.36	5

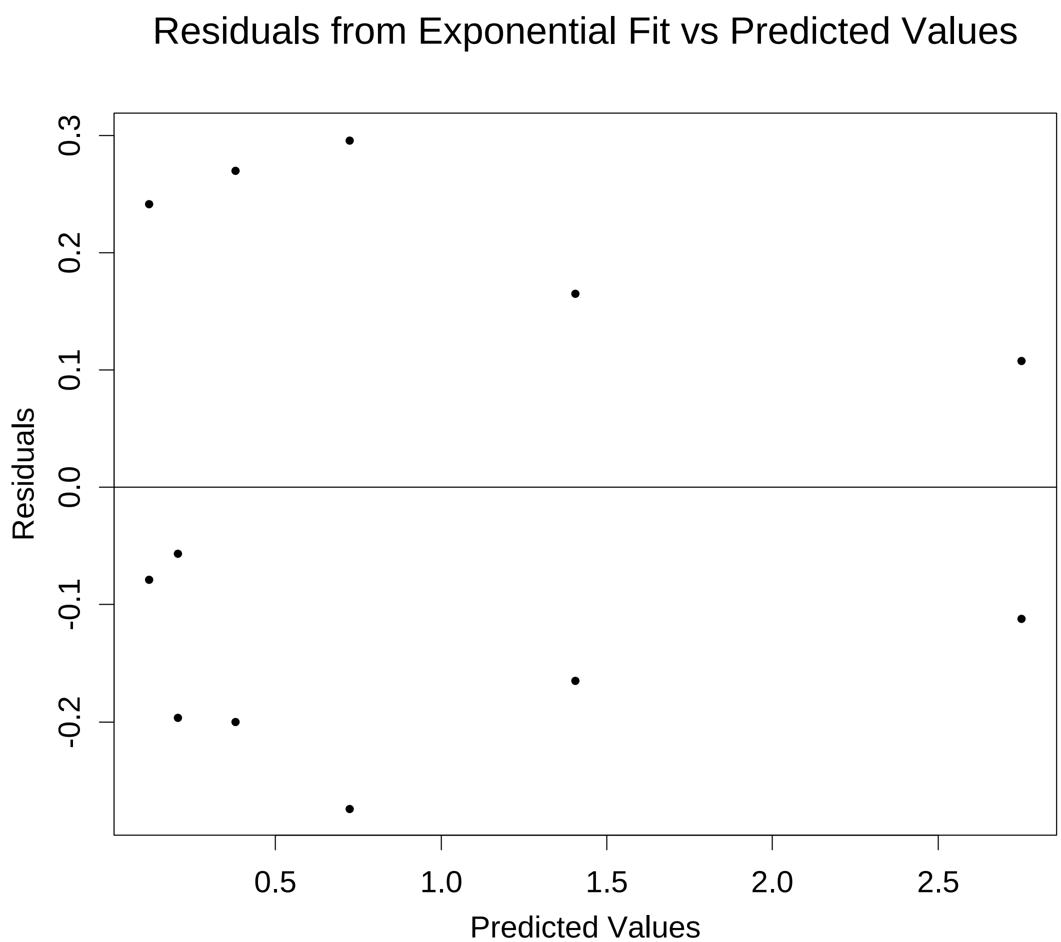
3a. Suppose scientific theory suggests that the optical readings should be related to the true concentrations by an exponential relationship of the form y = b1+b2*exp(-b3*x). Fit the theoretical model to the data and validate the quality of the fit.

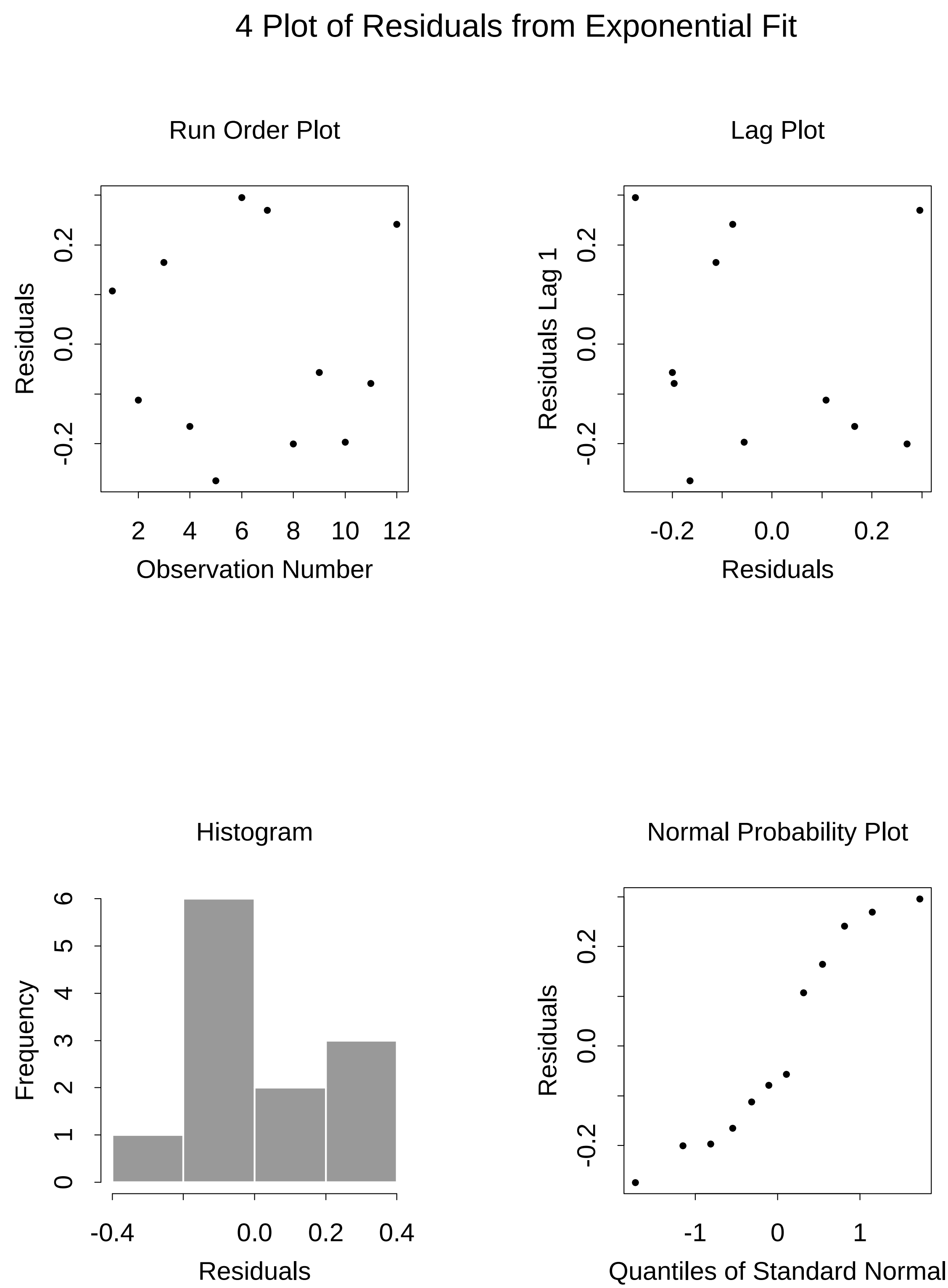
3b. Compute a 95% confidence interval for the true concentration of the chemical when the optical method gives a reading of 2.13 units.


Optical Measurement System Calibration Data
Optical Reading
0.0
0.5
1.0
1.5
2.0
2.5
0
1
2
3
4
5
True Concentration, mg/l


Optical Measurement System Calibration Data with Exponential Fit
Optical Reading
0.0
0.5
1.0
1.5
2.0
2.5
0
1
2
3
4
5
True Concentration, mg/l


Residuals from Exponential Fit vs True Concentration
Residuals
0.3
0.2
0.1
0.0
-0.1
-0.2
0
1
2
3
4
5
True Concentration, mg/l


Residuals from Exponential Fit vs Predicted Values
Residuals
0.3
0.2
0.1
0.0
-0.1
-0.2
0.5
1.0
1.5
2.0
2.5
Predicted Values



4 Plot of Residuals from Exponential Fit
Run Order Plot
Residuals
-0.2
0.0
0.2
2
4
6
8
10
12
Observation Number
Lag Plot
Residuals Lag 1
-0.2
0.0
0.2
-0.2
0.0
0.2
Residuals
Histogram
Frequency
0
1
2
3
4
5
6
-0.4
0.0
0.2
0.4
Residuals
Normal Probability Plot
Residuals
-0.2
0.0
0.2
-1
0
1
Quantiles of Standard Normal

Approximate Calibration Intervals by Inversion of Prediction Intervals

The basic steps to this approach of getting a calibration interval are:

1. solve the equation
$$f(x^*; \hat{\beta}_1, \ldots, \hat{\beta}_p) - y^* = 0$$
to get an estimate of x^*,
2. solve the equation
$$f(x^*; \hat{\beta}_1, \ldots, \hat{\beta}_p) + U(x^*) - y^* = 0$$
to get the lower confidence bound for x^*,
3. solve the equation
$$f(x^*; \hat{\beta}_1, \ldots, \hat{\beta}_p) - U(x^*) - y^* = 0$$
to get the upper confidence bound for x^*,

where $U(x^*) = t_{1-\alpha/2,n-p}\sqrt{s^2 + \vec{d}^{*T} V \vec{d}^*}$ as given on p. 233.

Computing the Estimate of x^*

The equation for estimating x^* for this problem is:

$$0.0287622+2.7232700\exp(-0.6827710\hat{x}^*)-2.13=0$$

Solving this equation numerically with generic nonlinear root-finding software yields the estimate $\hat{x}^* = 0.3798$

This equation can actually be solved analytically, but the that is not typically easy to do for most nonlinear models. A numeric approach is usually necessary for the computation of the confidence bounds too.

Computing $U(x^*)$

$$d^* = \begin{pmatrix} 0.0287622 \\ \\ \exp(-0.682771\hat{x}^*) \\ \\ -(2.72327)\hat{x}^*\exp(-0.682771\hat{x}^*) \end{pmatrix}$$

V can be obtained from the output from the fit by scaling the correlation matrix using the standard deviations of the estimated parameters:

```
      1.000*0.172*0.172 -0.678*0.172*0.211  0.846*0.172*0.142
V =  -0.678*0.172*0.211  1.000*0.211*0.211 -0.395*0.211*0.142
      0.846*0.172*0.142 -0.395*0.211*0.142  1.000*0.142*0.142
```

Computing $U(x^*)$

Doing the matrix multiplication symbolically if necessary and plugging these pieces into equations 2 and 3 on page 311 and solving gives the upper and lower confidence bounds for the calibration interval.

For this problem the interval turns out to be:

$$(-0.0384, 0.3798, 1.0436)$$